Introduction to ST 2022-6 & ST 2110

by Renaud Lavoie

Introduction

This paper aims to describe and explain the differences between the SMPTE standards ST 2022-6 and ST 2110, weigh the pros and cons of each of those two standards, and carve out some specific use cases.

Since the debut of IP streaming usage for broadcasting workflows, several solutions have emerged from different origins, the earliest and most popular standard being ST 2022, introduced in 2007. While the first parts of the ST 2022 standard cover IP protocols for compressed video signals, the ST 2022-6 part published in 2012 established a method for carrying uncompressed video.

Some people might say that ST 2022-6 is a thing of the past that is using too much bandwidth in your network, but there are several use cases that can truly benefit from the simplicity of ST 2022-6 workflows. ST 2110, on the other hand, is the newcomer in the world of video/audio/ancillary over IP, but quickly became the focus of the broadcast industry's efforts. The clear goal of ST 2110 is to achieve IP production workflows that are even more efficient than traditional SDI workflows.

For more information, have a look at Renaud talks IP - 2110 vs 2022-6 on YouTube.

1. SMPTE ST 2022-6

The move from SDI to IP was done step by step, and the first one indeed was ST 2022-6. The concept was pretty simple: taking all of the SDI information, including vertical and horizontal blanking, and packetizing it for the IP network. As simple as that. The following picture illustrates the principle:

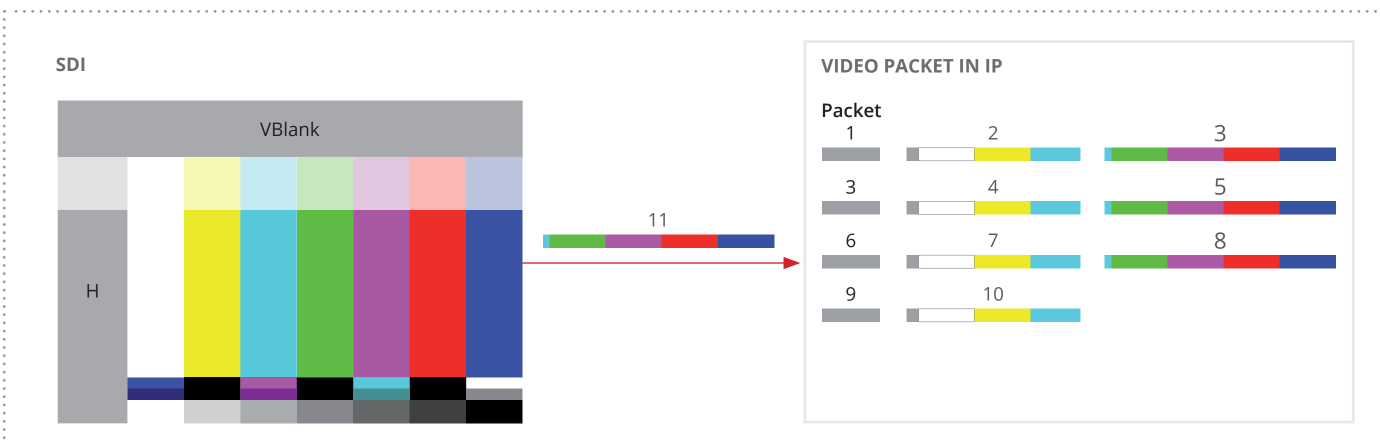

Figure 1. ST 2022-6 frame to packets

Note: The typical maximum transmission unit (MTU or maximum amount of data in the packet) is approximately 1500 bytes of data over ethernet (excluding the IP headers). ST 2022-6 specifies the size of the packet to ensure compatibility between vendors:

> *The video luminance and color-difference values shall be encapsulated into 1376 octet media payloads. The last datagram of the video frame, being only partially filled with luminance and color-difference values, shall have additional null octets added to achieve a total length of 1376 octets. This is not considered padding at the RTP layer, therefore the padding bit iin the RTP header shall be set to zero.*

Figure 2. Extract from SMPTE ST 2022-6 specification

As you can see in the extract, only the last packet (datagram) shall have null data to achieve 1376 octets. Therefore, all the packets have the same size.

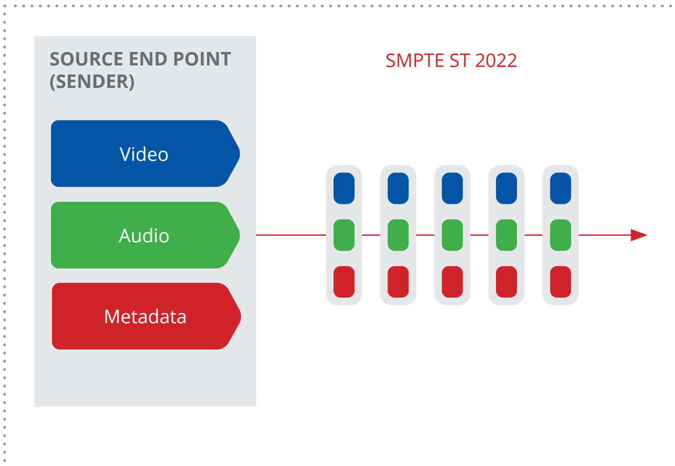

Figure 3. ST 2022-6 packets generation

In ST 2022-6, the audio, video and ancillary share the same flow (unicast or multicast stream). Because ST 2022-6 is only packetizing SDI and encapsulating that entire SDI signal in IP packets, the total bandwidth of the SDI signal is sent over the network, with the addition of ethernet headers which add approximately 3% on top of the total SDI bandwidth.

Another drawback of ST 2022-6 is the fact that, because all the essences are together, SDI audio embedders & de-embedders are still needed to do audio mixing, shuffling, etc. The only optimization to the system is the removal of the “Distribution Amplifier (DA)”, which is covered in the equipment list by the ability to transmit one signal to multiple receivers via multicast in the IP network.
One last important point is that because ST 2022-6 is more of a transport method, SMPTE did not specify any synchronization method in ST2022-6. As all the essences are still together, you need to go back to SDI to manipulate the video, audio and ANC.

Takeaway: ST 2022-6 is simply mapping SDI over IP.

2. SMPTE ST 2110

ST 2110 goes back to a time where essences were separated, to the early days of broadcast. In the analog world, video, audio and ANC were separated, but with digitalization and serialization, it made sense to combine the essences in SDI (Serial Digital Interface). With the move to IP, this constraint fades away and once again, separating the essence makes more sense. Think for a second about the audio control room: Why send large quantities of video streams when you could just send few and only send lots of audio streams (AES67 in ST 2110). The following picture shows the video packets in IP. Note that blanking does not exist anymore in ST 2110.

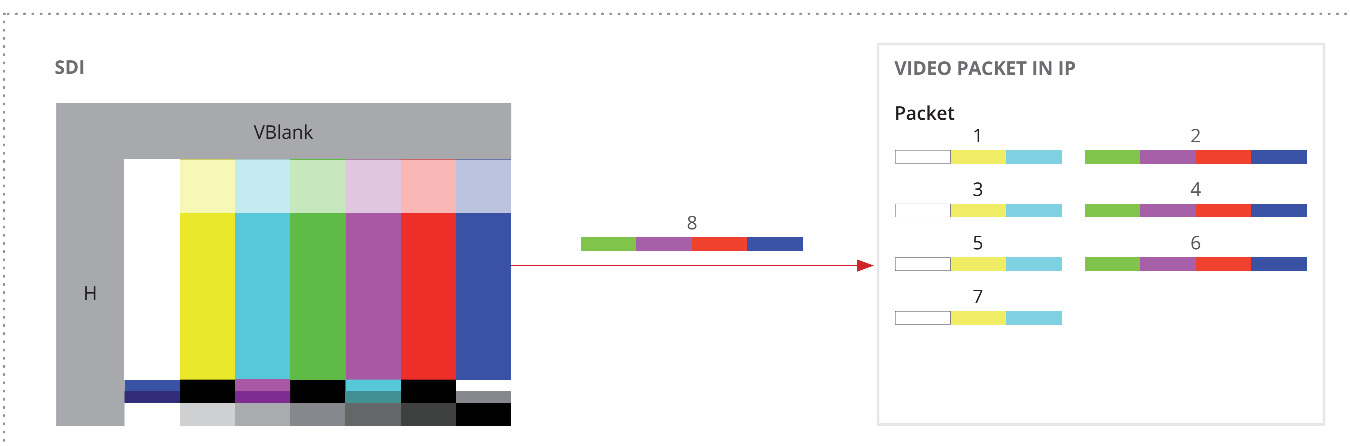

Figure 4. ST2110 frame to packets

Video

The video essence defined first in ST 2110 part 20 (ST 2110-20) is based on RFC4175, which defines a flexible packetization scheme for encapsulating uncompressed video up to 32K pixels:

> *This memo specifies a packetization scheme for encapsulating uncompressed video into payload format for the Real-time Transport Protocol, RTP. It supports a range of standard- and high-definition video formats, including common television formats such as ITU BT.601, and standards from the Society of Motion Picture and Television Engineers (SMPTE), such as SMPTE 274M and SMPTE 296M. The format is designed to be applicable and extensible to new video formats as they are developed.*

Figure 5. Abstract of RFC4175

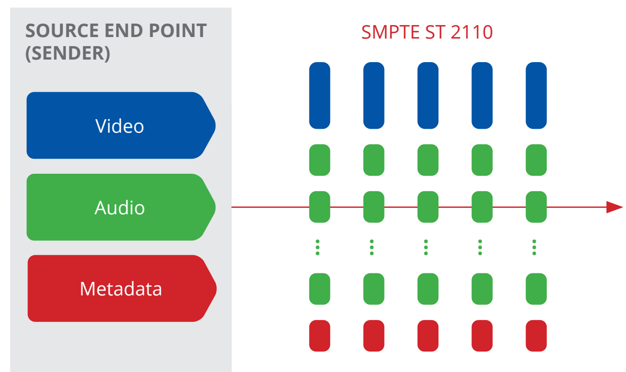

Figure 6. ST 2110 packets generation

Other parts for video have been added to the specification. At the time of writing this document:
• ST 2110-21: addressing traffic shaping and delivery timing of uncompressed video

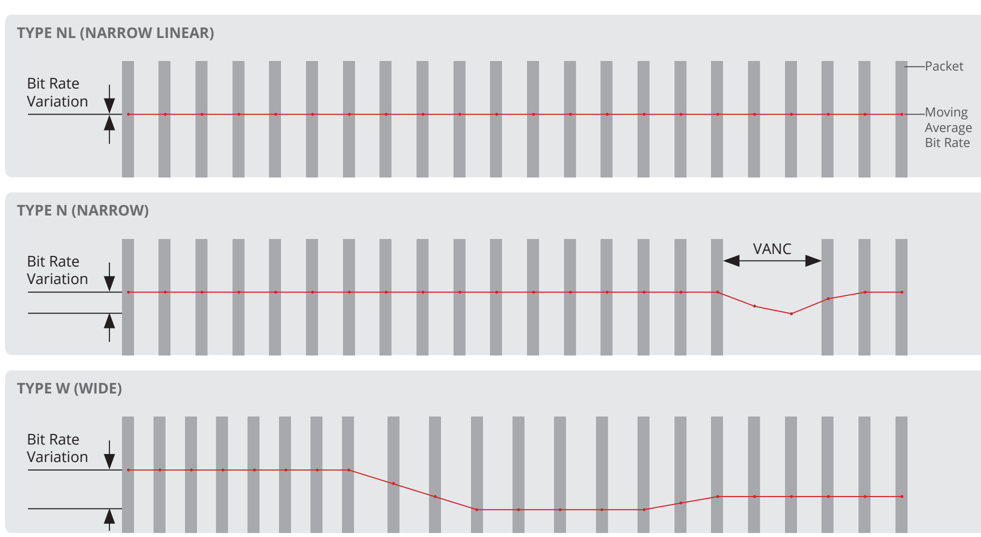

Figure 7. Different types of ST 2110 senders defined in ST 2110-21

For more information, please see [Renaud talks IP – Wide versus Narrow Sender] on YouTube.

- ST 2110-22: Compressed video (see next section)
- ST 2110-23: Recommended practice on a methodology to split high bandwidth single video streams, lower bandwidth streams can be 2SI and square division

This document proposes a methodology to (i) split high bandwidth single video essence streams into several lower bandwidth SMPTE ST 2110-20 tributary streams, and (ii) to describe the appropriate grouping and signaling of these multiple SMPTE ST 2110-20 streams (SDP declarations, addressing conventions, RTP time stamp constraints, ...). This RP does not intend to define new decomposition methods, but refers to existing approaches. Lower bandwidth streams can be made according to
- The 2SI and Square Division (SD) decomposition mechanisms for UHD content
- Temporally decomposing the video into lower framerate streams

Figure 8. SMPTE RP 2110-23:2019

Compressed video (ST 2110-22)

ST 2110-22 specifies the payload for constant bite rate compression video. One popular choice is JPEG-XS over ST2110. The -22 standard defines rasters of up to 32K lines per 32K pixels. It also specifies that the payload (compressed signal) can have any color sampling from 4:1:1 to 4:4:4 and that the frame rate and bit depth (up to 16bits) can be flexible. It supports the commonly used HDR formats PQ and HLG.

Audio (ST 2110-30, ST 2110-31)

ST 2110-30 is based on the AES67 format and is used to transport uncompressed audio in PCM format.
ST 2110-31 specifies the real-time RTP-based transport of AES3, which can be transparent or compressed audio over AES3 format.
There are many use cases where it makes sense to separate audio from other essences, like the aforementioned audio control room. Here, separate audio streams enable the use of lower-cost and lower-bandwidth switches. These bandwidth savings allow for a general simplification of the network.

Ancillary (ST 2110-40)

ST 2110-40 uses RFC 8331 to indicate how to wrap ancillary in IP. The goal of separating the ancillary data is to remove the need for VANC inserters. In SDI, closed caption, teletext and any other VANC data need to be inserted by VANC inserters, which sometimes results in large number of additional devices. ST 2110-40, with the ability to keep ancillary isolated, can really help reduce system complexity.

The following picture shows a captioner node responsible for generating the closed captioning in ST 2110. The triggers are usually used for ads insertion and other triggers used in ancillary.

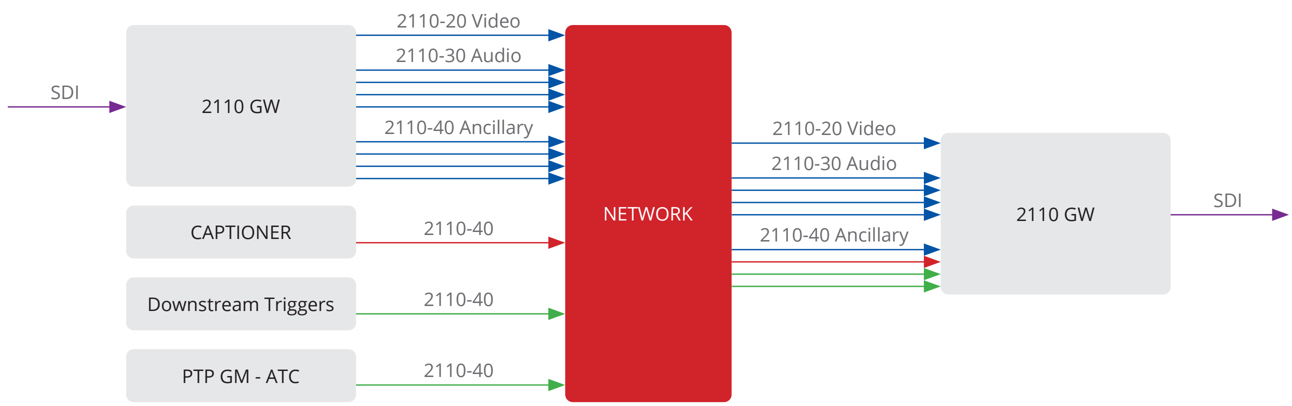

Figure 9. Closed capturing generation in ST 2110

Another key advantage is to gain speed and simplicity if a piece of equipment only manipulates ST 2110-40, for example a caption processor.

Other parts for ancillary are being added to the specification. At the time of writing this document:
- ST 2110-41: Fast metadata, allowing metadata to be sent faster than ST 2110-40 for critical information, such as ads insertion.

System Timing and Definitions (ST 2110-10)

SMPTE ST 2110-10 is a family of documents defining an extensible system of RTP-based essence streams referenced to a common reference clock. The RTP is based on the RFC3550 protocol and shall conform to the RTP profile in RFC3551. ST 2110-10 also defines the system timing model and synchronization based on the precision time protocol (PTP), IEEE 1588-2008. If the device is in slave state, the device internal clock shall be synchronized to the common reference clock. The two SMPTE specifications on the PTP are:

ST 2059-1: PTP epoch
The media clock shall have a value of zero at the SMPTE Epoch as specified in SMPTE ST 2059-1.

ST 2059-2: PTP profile
All devices conforming to ST 2110 need to support the SMPTE ST 2059-2 PTP profile. From the SMPTE ST 2110-10 specification, the parameter range inside ST 2059-2 is constrained to support limits of AES67:2015 to enable equipment to interchange audio streams with AES67 compliant devices. The AES has issued AES-R16-2016, a technical report regarding compatibility of parameters between AES67 media profile and SMPTE ST 2059-2.

Takeaway: ST 2110 is separating the essences over IP.

3. Use Cases

ST 2022-6 transports all video, audio and ancillary data essences in a single stream. This is not practical when audio essences must be used in different processes in the system, but can work nicely for playout and transport. Keeping the essences together in a single stream allows users not have to worry about multiple flows an different latencies between essences. This transport method also ensure that the entire package arrives at the destination with all essences in time with each other, just like SDI

ST 2110 provides individual transport of essences, which comes with the benefit of using audio essences for audio specific devices without having to transport the video and ancillary data. This removes the need for a device to extract the audio essence from the stream, allowing cost savings on network links and more flexible production setups with essences only sent to the destinations on the network where they are required

4. Conclusion

We hope that this paper provided a useful overview of the basic functionalities of two of the most common IP video standards. After giving a quick rundown of the basics of ST 2022-6, the all-included SDI over IP standard, we explored the parts that compose the production-oriented ST 2110 and some of its particularities, such as synchronization, separation of the essences, and different types of senders.

In general, ST 2022-6 and ST 2110 don’t have a lot in common, except the fact that both are SMPTE standards for video, audio and ancillary over IP. While ST 2022-6 is more of a transport solution, i.e. for playout and transport over long distance, ST 2110 is useful for production environments where you or your customers need to shuffle audio, change ancillary, compress video… all the regular steps to do a production!

For more information and details about these two standards, have a look at our youtube playlist Renaud talks IP or visit riedel.net to find new articles and whitepapers.